EXTREMES WITH COVARIATE EFFECTS - PRACTICALITIES AND PROBLEMS

Philip Jonathan
Shell Technology Centre Thornton, Chester, UK.
philip.jonathan@shell.com

Kevin Ewans
Shell International Exploration and Production, Rijswijk, NL.
kevin.ewans@shell.com

24 September 2008

Overview

Motivation

- Regulatory requirements ad-hoc (if not inconsistent) w.r.t. accommodation of covariate effects and estimation of (e.g.) directional and seasonal design values for coastal and ocean structures.
- Statistics literature provides a framework for rational and consistent estimation, but many issues unresolved.

This talk

- Analysis procedure with application to North Sea design
- Some pressing issues in modelling and interpretation

Procedure in a nut shell: I

- Hindcast data for multiple locations in neighbourhood. Extract *storm peaks* (to eliminate temporal dependence) over threshold u.
- Assume extremal characteristics of all locations marginally identical, although dependent. Goal is to estimate distribution of n-year return value q_n for any single location.
- Estimation using NHPP: storm arrival rate μ, GP shape γ and scale σ.
- Accommodate covariate effects: μ, γ, σ and u vary with covariates (e.g. direction, season, time). u estimated before hand as high (local) quantile (sensitivity to threshold choice).
- Maximise likelihood, penalised by parameter roughness w.r.t. covariates. Diagnostics for model fit. Cross-validation for optimal roughness. Block bootstrap for parameter uncertainty pointwise.

Procedure in a nut shell: II

- Simulate to estimate properties of q_n.
- Estimate q_n also for partitions w.r.t. covariates. Estimate and accommodate storm dissipation effects.
- Present findings in engineering terms.

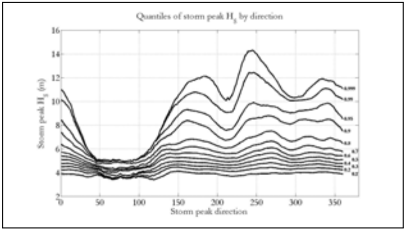
Procedure published here:

- Effect of combining locations on estimation uncertainty (Jonathan and Ewans 2007b).
- Illustrations of extent of covariate effects on extreme quantile estimates (Jonathan et al. 2008).
- Modelling directional extremes in the Gulf of Mexico and Northern North Sea (Jonathan and Ewans 2007a, Ewans and Jonathan 2008).

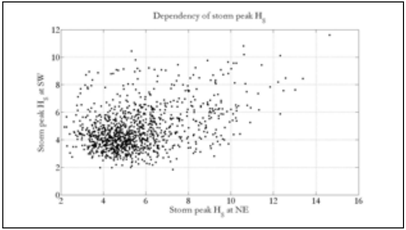
Data and pre-processing

- Significant wave height H_S values from a Northern North Sea hindcast, for October 1964 to September 1998 inclusive, at 3 hour intervals.
- For approximate location of 2° longitude, 61° latitude, selected 100 grid points on 10 x 10 rectangular lattice covering an area of approximately 5° longitude, 3° latitude centred at the location of interest.
- For each storm period for each grid point, isolated storm peak significant wave height, H_S^{sp}, and corresponding wave direction, θ.
- Estimated threshold for extreme value analysis based on local quantile (with direction)
- Estimated directional dissipation for a storm (w.r.t. storm peak H_S, with direction)

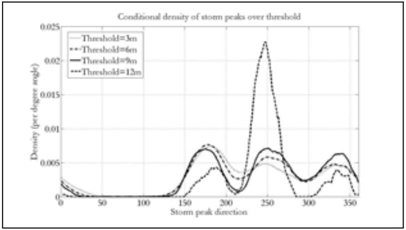
Exploratory analysis



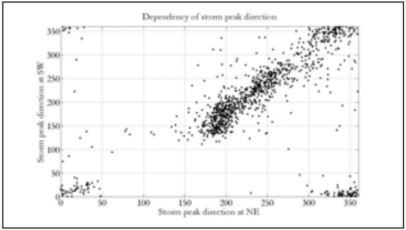
Quantiles of $H_S^{sp}(\theta)$



H_S^{sp} for NE and SW

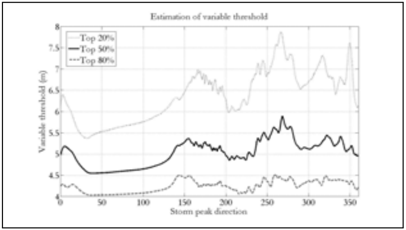


Conditional densities of $H_S^{sp}(\theta)$



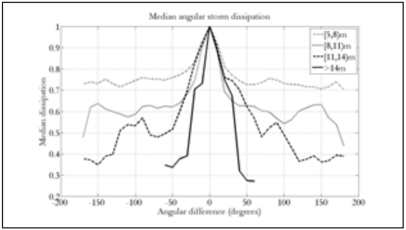
θ for NE and SW

Threshold and dissipation



Threshold $u(\theta)$

- Given interval Θ of θ, threshold estimated empirically as quantile of $H_S^{sp}(\Theta)$
- Local median used for modelling henceforth



Dissipation $\rho(\Delta\theta)$

- For a given storm event, $\rho(\Delta\theta) = H_S(\Delta\theta)/H_S^{sp}$
- Dissipation quantifies influence of storm on extremes in directions other than θ

Generalised Pareto Modelling: I

Given $\{X_i\}_{i=1}^n$, $\{\theta_i\}_{i=1}^n$, distribution of storm peaks above variable threshold $u(\theta)$ assumed GP with cdf $F_{X_i|\theta_i,u}$:

$$
\begin{aligned}
F_{X_i|\theta_i,u}(x) &= P\left(X_i \leq x|\theta_i, u(\theta_i)\right) \\
&= 1 - \left(1 + \frac{\gamma(\theta_i)}{\sigma(\theta_i)}\left(x - u(\theta_i)\right)\right)_+^{-\frac{1}{\gamma(\theta_i)}}
\end{aligned}
$$

γ and σ vary smoothly with θ, assumed to follow Fourier form $\sum_{k=0}^{p}\sum_{b=1}^{2} A_{abk}t_b(k\theta)$.

Generalised Pareto Modelling: II

Penalised negative log likelihood is l^*:

$$l^* = \sum_{i=1}^{n} l_i + \lambda \left(R_\gamma + \frac{1}{w} R_\sigma \right)$$

Unpenalised negative log likelihood is:

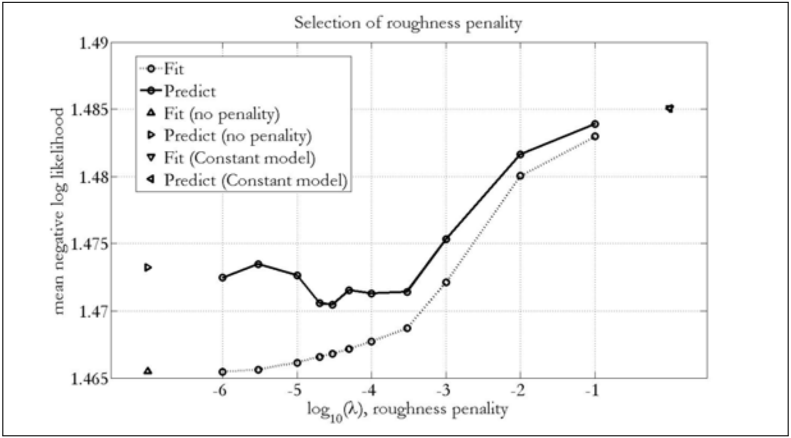
$$l_i = \log \sigma(\theta_i) + \left(\frac{1}{\gamma(\theta_i)} + 1 \right) \log \left(1 + \frac{\gamma(\theta_i)}{\sigma(\theta_i)} (X_i - u(\theta_i)) \right)_+$$

Roughness of γ is given by:

$$R_\gamma = \int_0^{2\pi} \left(\frac{\partial^2 \gamma}{\partial \theta^2} \right)^2 d\theta = \sum_{k=1}^{p} \pi k^4 \left(\sum_{b=1}^{2} A_{1bk}^2 \right)$$

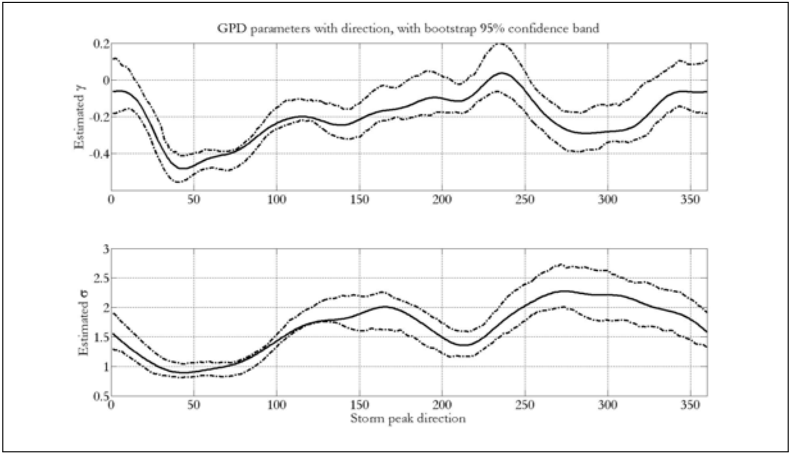
Analogous expression for roughness of σ

Cross-validation for roughness



Optimal λ

Forms of γ and σ with block bootstrap



$\gamma(\theta), \sigma(\theta)$, with block bootstrap 95% confidence interval

Poisson Modelling: I

Non-homogeneous Poisson process model. The negative log-likelihood written:

$$l(\mu,\gamma,\sigma) = l_N(\mu) + l_W(\gamma,\sigma)$$

where l_N is the (negative) log-density of the total number of exceedances (with rate argument μ), and l_W is the (negative)log-conditional-density of exceedances given a known total number N). Inferences on μ made separately from those on γ and σ.
The Poisson process log-likelihood, for arrivals at times $\{t_i\}_{i=1}^n$ in period P_0 is:

$$l_N(\mu) = -(\sum_{i=1}^{n} \log \mu(t_i) - \int_{P_0} \mu(t)dt)$$

Poisson Modelling: II

Or approximately (Chavez-Demoulin and Davison 2005):

$$\hat{l}_N(\mu) = -(\sum_{j=1}^{m} c_j \log \mu(j\delta) - \delta \sum_{j=1}^{m} \mu(j\delta))$$

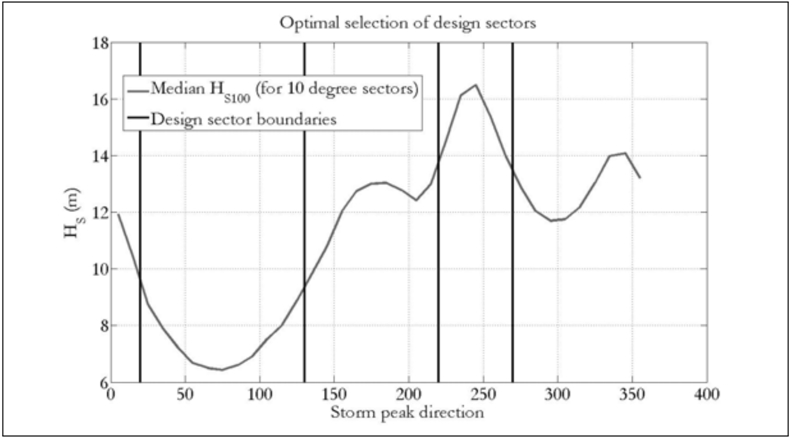
where $\{c_j\}_{j=1}^{m}$ is the number of occurrences in each of the m sub-intervals. W

We estimate storm occurrence rate adopting a Fourier form for Poisson intensity μ as a function of θ, penalising its roughness R_μ:

$$\hat{l}_N^*(\mu) = \hat{l}_N(\mu) + \kappa R_\mu$$

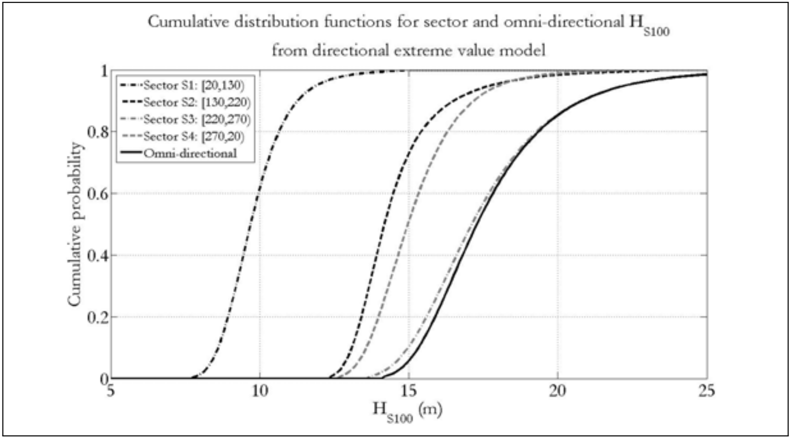
R_μ has form analogous to that of R_γ or R_σ. Cross-validation and block bootstrapping used similarly.

Defining appropriate directional sectors



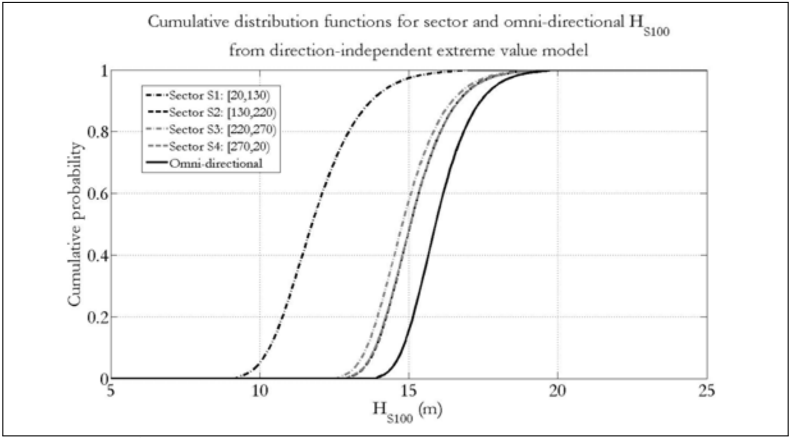
Optimal boundaries to reduce within-sector variability

100-year storm peak cdf using directional γ,σ



H^{sp}_{S100} for directional sectors and omni-directionally

100-year storm peak cdf using constant γ,σ



H^{sp}_{S100} for directional sectors and omni-directionally

Design values for different design strategies

- Specification of an omni-directional non-exceedance probability does not uniquely define *sector* design values.
- Different strategies possible:
  - Design to omnidirectional H_{S100}^{sp}
  - Design to equal sector non-exceedence probability
  - Design to optimise a specified cost function

Sector	Angle	Risk-Cost Optimal			Omni-directional			Equal Non-exceedence probability		
Directional model		RiskCost	Value (m)	Quantile	RiskCost	Value (m)	Quantile	RiskCost	Value (m)	Quantile
S1	[20,130)	8.6	14.0	0.99	12.3	17.5	1.00	9.2	10.9	0.85
S2	[130,220)		15.6	0.82		17.5	0.94		15.8	0.85
S3	[220,270)		18.6	0.73		17.5	0.57		19.8	0.84
S4	[270,20)		16.6	0.84		17.5	0.92		16.6	0.84
Constant model		RiskCost	Value (m)	Quantile	RiskCost	Value (m)	Quantile	RiskCost	Value (m)	Quantile
S1	[20,130)	7.7	14.0	0.92	10.2	16.0	0.99	7.9	13.3	0.85
S2	[130,220)		16.2	0.82		16.0	0.79		16.3	0.85
S3	[220,270)		15.9	0.81		16.0	0.83		16.0	0.84
S4	[270,20)		16.2	0.82		16.0	0.79		16.3	0.84

Issues

Generic

- Sample size (c.f. estimates required for long return periods)
- Measurement (or hindcast) uncertainty (especially for extreme values)
- Temporal dependence ($\therefore$ "storm peak" analysis)
- Spatial dependence ($\therefore$ block bootstrap)
- Model form (e.g. GP versus Weibull ...) and complexity
- Transformation of variables (e.g. weighting locally w.r.t. covariate)
- Combination of variables (e.g. joint modelling, *structural response - based* analysis)

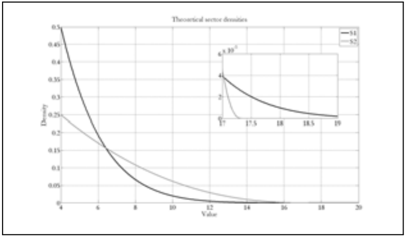
Specific

- Reflecting model specification and fitting uncertainty in design values
  - Threshold selection
  - Model stiffness
  - Dissipation
- Specification and interpretation of design conditions in engineering context

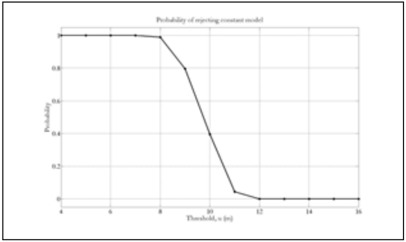
Model selection - a toy problem

- Two homogeneous directional sectors S_1 and S_2, extremes are GP-distributed
- γ, σ and u values potentially different between sectors
- Random sample size 1250 from each sector corresponding to 25 years
- Test
  - H_0: $\gamma_1 = \gamma_2$, $\sigma_1 = \sigma_2$
  - H_A: $\gamma_1 \neq \gamma_2$, $\sigma_1 \neq \sigma_2$
- Cases (γ_1,σ_1,u_1) and (γ_2,σ_2,u_2)
  - Case 1 (-0.1,2,4) and (-0.3,4,4)
  - Case 2 (-0.1,3,4) and (-0.3,4,6)

Case 1 - Model selection

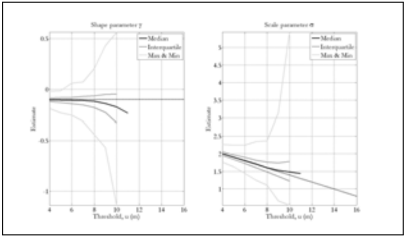


Sector densities (theory)

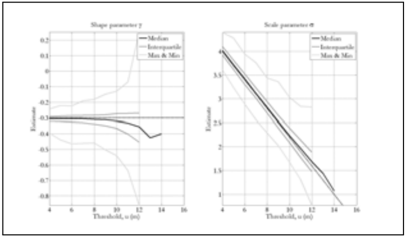


Probability of rejecting H_0

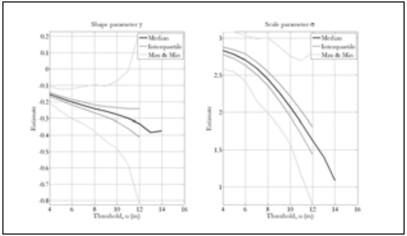
Case 1 - Parameter estimates with threshold



H_A: S_1 estimates with threshold

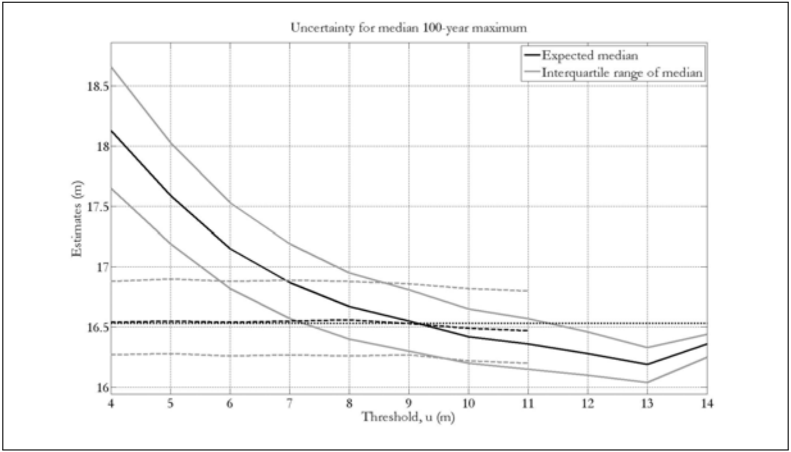


H_A: S_2 estimates with threshold



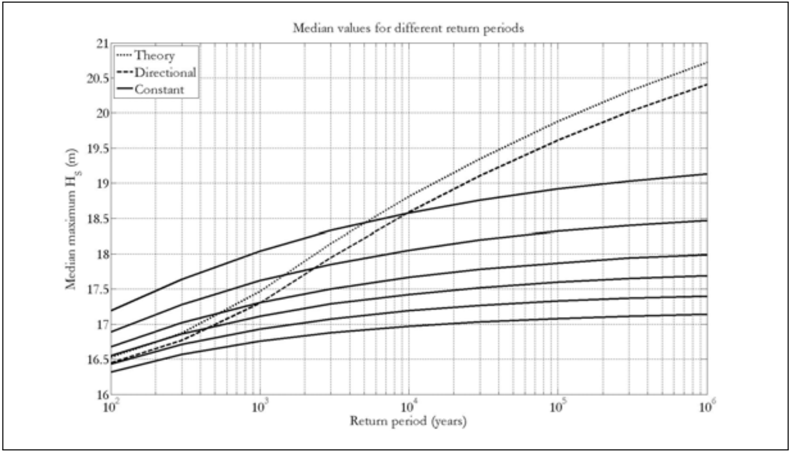
H_0: estimates with threshold

Case 1 - Estimated median 100-year maximum



Median 100-year event: true (dot), H_A (dash), H_0 (full)

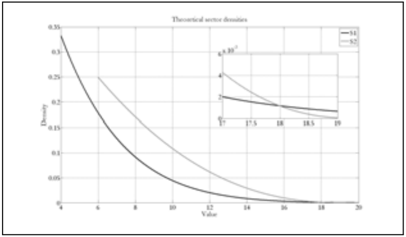
Case 1 - Implications for long return periods



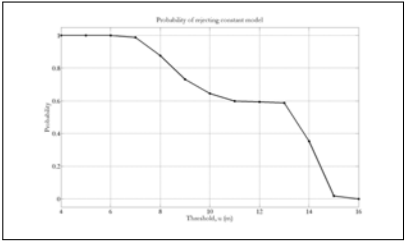
Median P-year event: true (dot), H_A (dash),

H_0 (full, for thresholds 6,7,...,11m)

Case 2 - Model selection

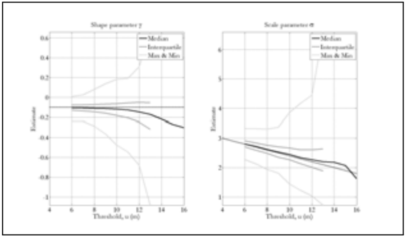


Sector densities (theory)

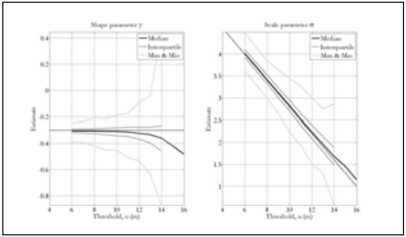


Probability of rejecting H_0

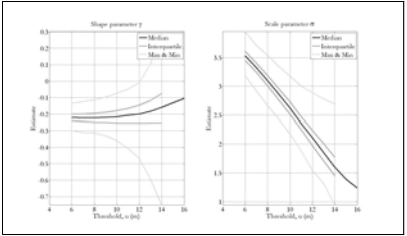
Case 2 - Parameter estimates with threshold



H_A: S_1 estimates with threshold

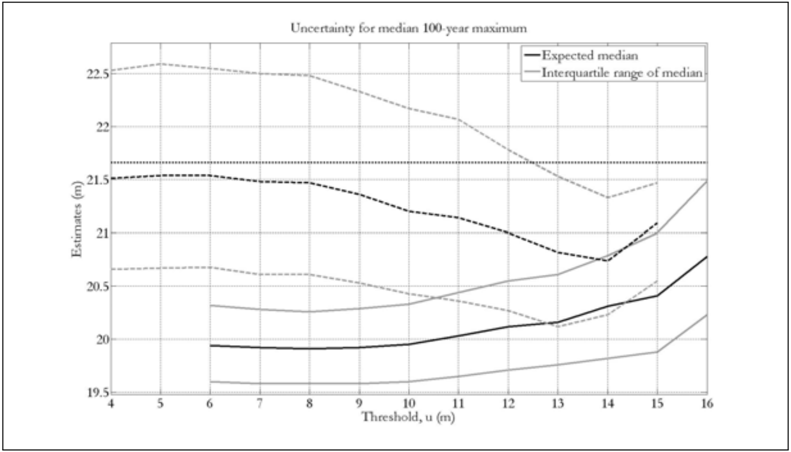


H_A: S_2 estimates with threshold



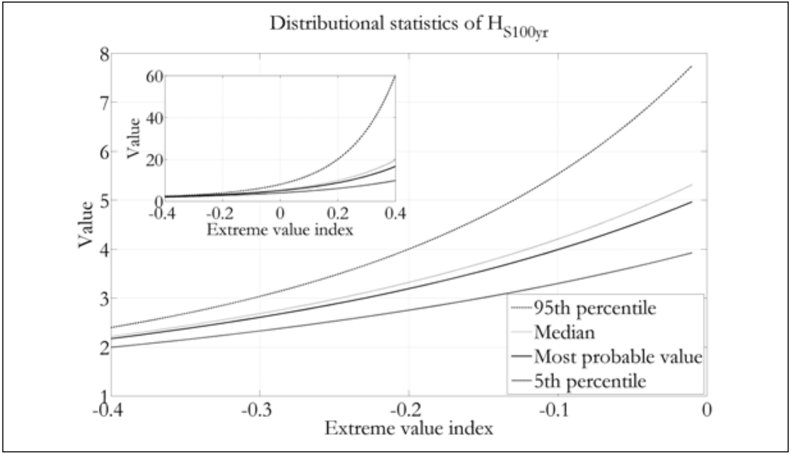
H_0: estimates with threshold

Case 2 - Estimating median 100-year maximum



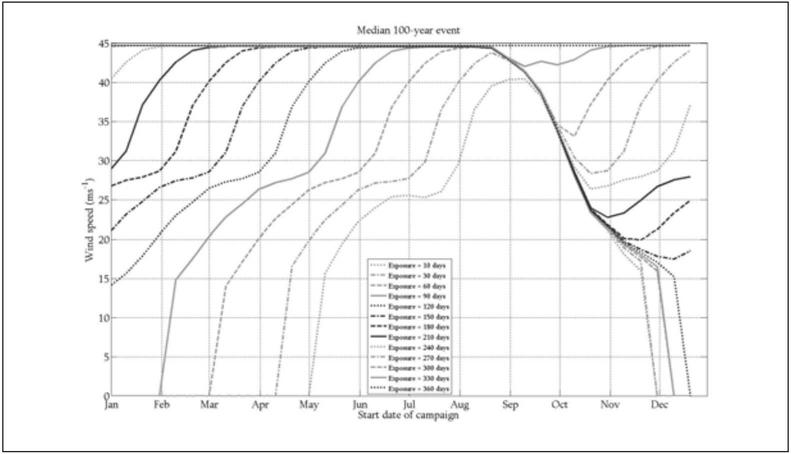
Median 100-year event: true (dot), H_A (dash), H_0 (full)

Interpretation: Ratios of extreme quantiles with γ



σ=1. Moments of H^{sp}_{S100} vary with γ

Interpretation: seasonal extremes



Design values for short-term deployments in Gulf of Mexico

Future work

- Model multiple covariate effects (e.g. more general smoothers).
- Model spatial and temporal dependence explicitly (e.g. extreme quantiles for region rather than single location).
- Improved modelling of dissipation effects.
- Jointly model multiple variables (wind, waves, current, e.g. Heffernan and Tawn 2004), compare inferences with *response-based* approaches.
- Extremes estimates incorporating uncertainties from model and threshold specification (e.g. predictive distributions).
- Models that better exploit the underlying physics (e.g. for hurricanes)
- Influence design practice. Regulators (e.g. API) currently reviewing methods for seasonal and directional design. Bridge industry and academia, communicate.

References

V. Chavez-Demoulin and A.C. Davison. Generalized additive modelling of sample extremes. *J. Roy. Statist. Soc. Series C: Applied Statistics*, 54: 207, 2005.

K. C. Ewans and P. Jonathan. The effect of directionality on Northern North Sea extreme wave design criteria. *J. Offshore Mechanics Arctic Engineering*, 2008. (Accepted October 2007. DOI:10.1115/1.2960859).

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B*, 66:497, 2004.

P. Jonathan and K. C. Ewans. The effect of directionality on extreme wave design criteria. *Ocean Engineering*, 34:1977–1994, 2007a.

P. Jonathan and K. C. Ewans. Uncertainties in extreme wave height estimates for hurricane dominated regions. *J. Offshore Mechanics Arctic Engineering*, 129:300–305, 2007b.

P. Jonathan, K. C. Ewans, and G. Z. Forristall. Statistical estimation of extreme ocean environments: The requirement for modelling directionality and other covariate effects. *Ocean Engineering*, 35: 1211–1225, 2008.

Thanks for listening.

philip.jonathan@shell.com